

---

# Boolean models for genetic regulatory networks

**Cynthia Olson Reichhardt**

*T-12*

*Los Alamos National Laboratory*

**Kevin Bassler**

*Department of Physics*

*University of Houston*

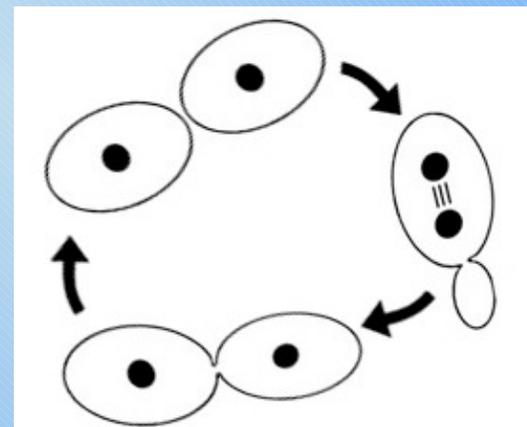
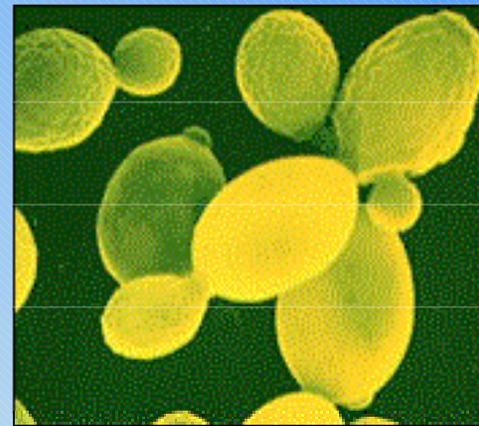
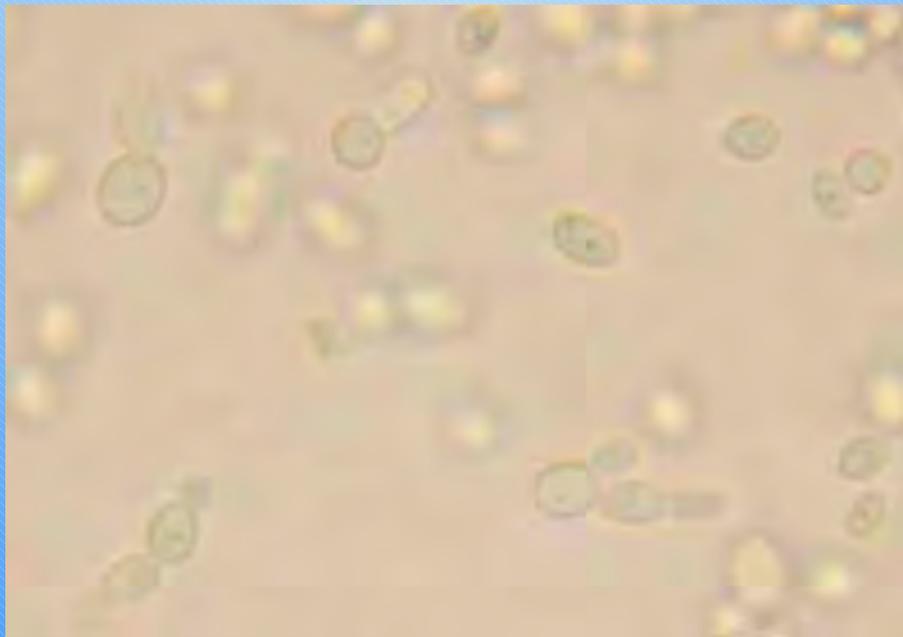


# Outline

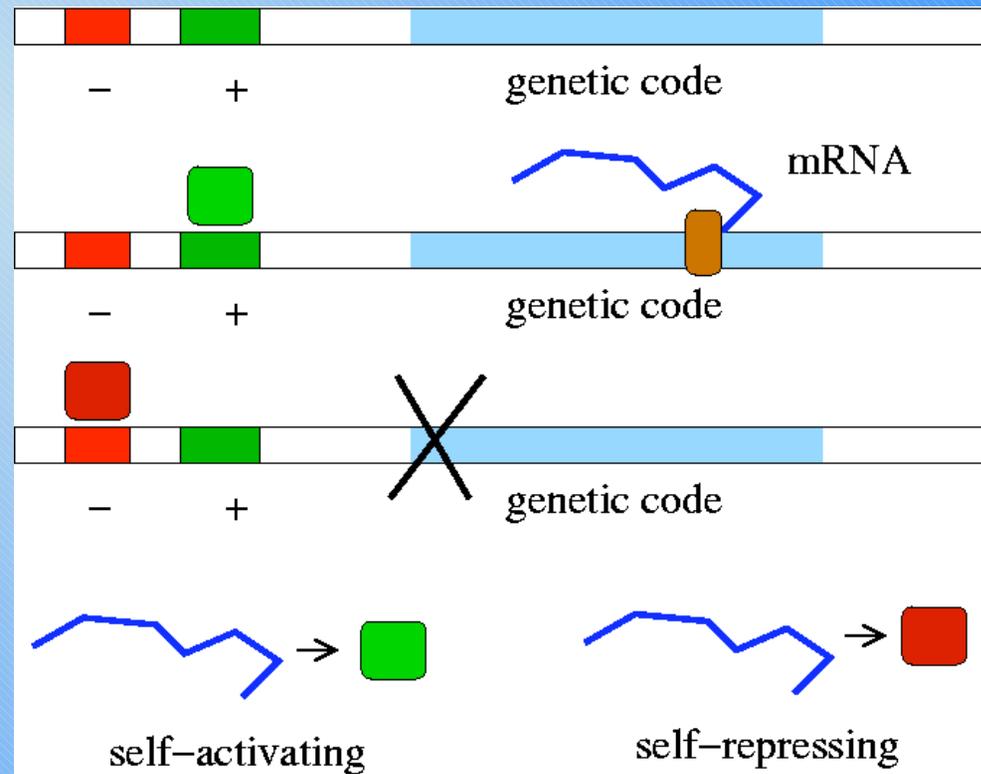
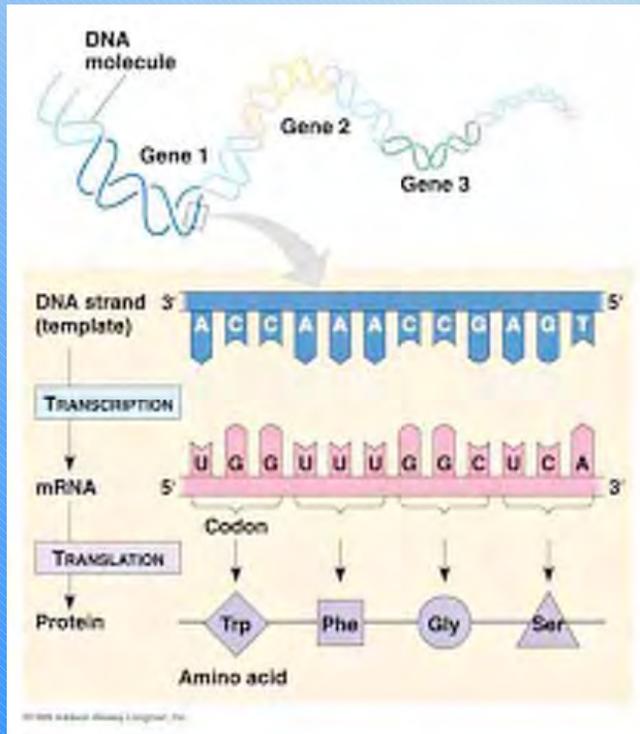
---

- Introduction: Cell cycles, genetic regulatory networks, and phenotypic development
- Phenotypes, mutations, and genetic variation. The canalization concept
- Representations of genetic logic: Boolean network models
- Cell type and cycle length in Kauffman network models
- Mapping of the gene strategy tables to Ising hypercubes
- Group theoretic concepts (brief review)
- Symmetry properties of strategies for combinatoric enumeration
- Preserving the phenotype through network interactions
- Summary

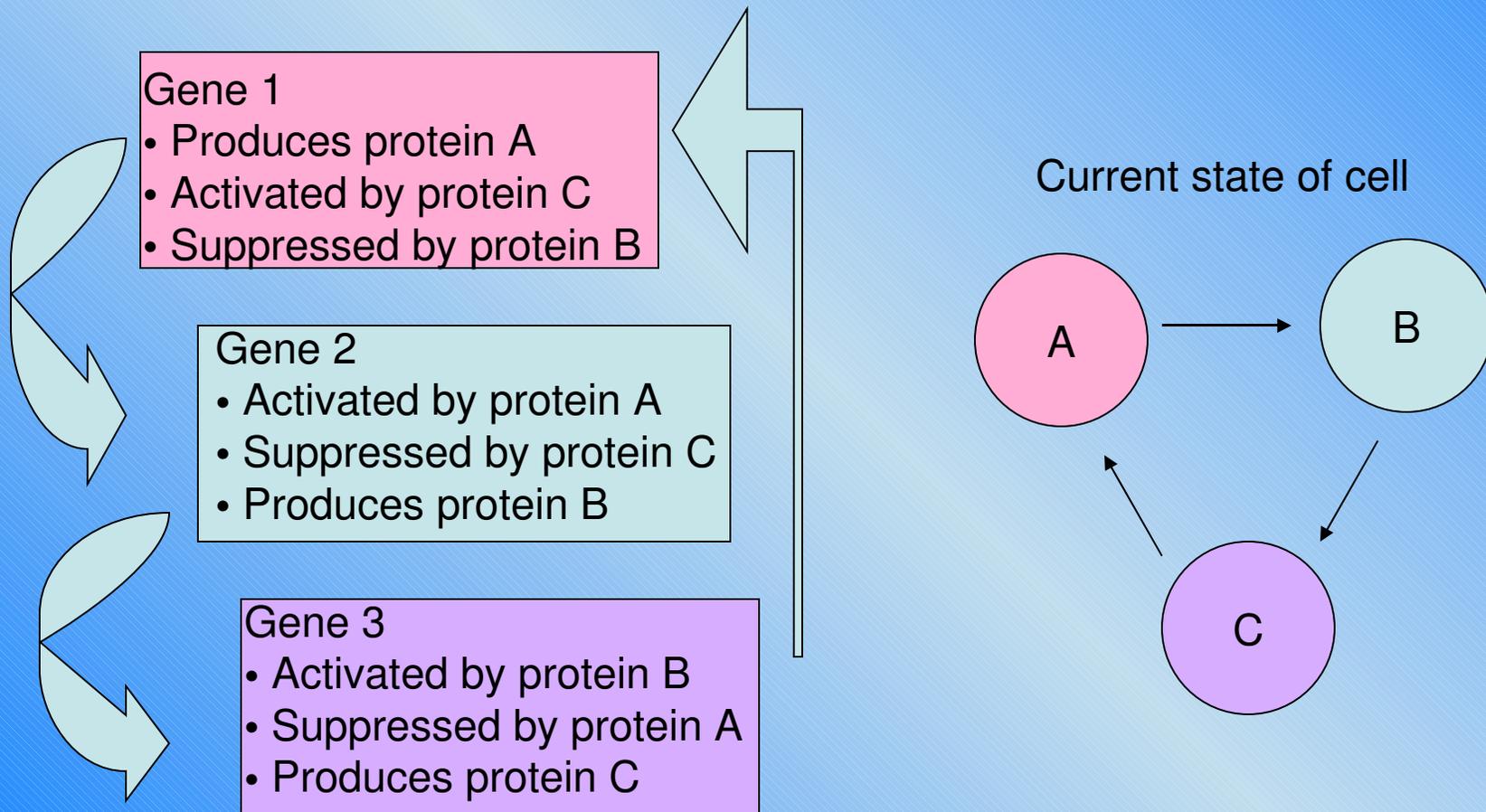
# Cell cycle in *Saccharomyces cerevisiae*



# The brain of yeast



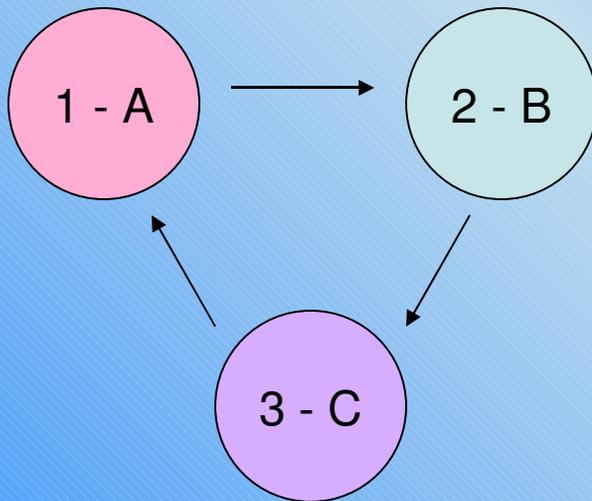
# Gene-gene interactions



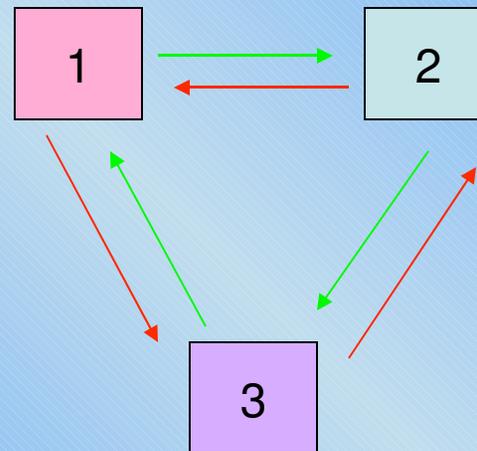
# Gene-gene interaction **network**

---

Cell **cycle** (temporal)

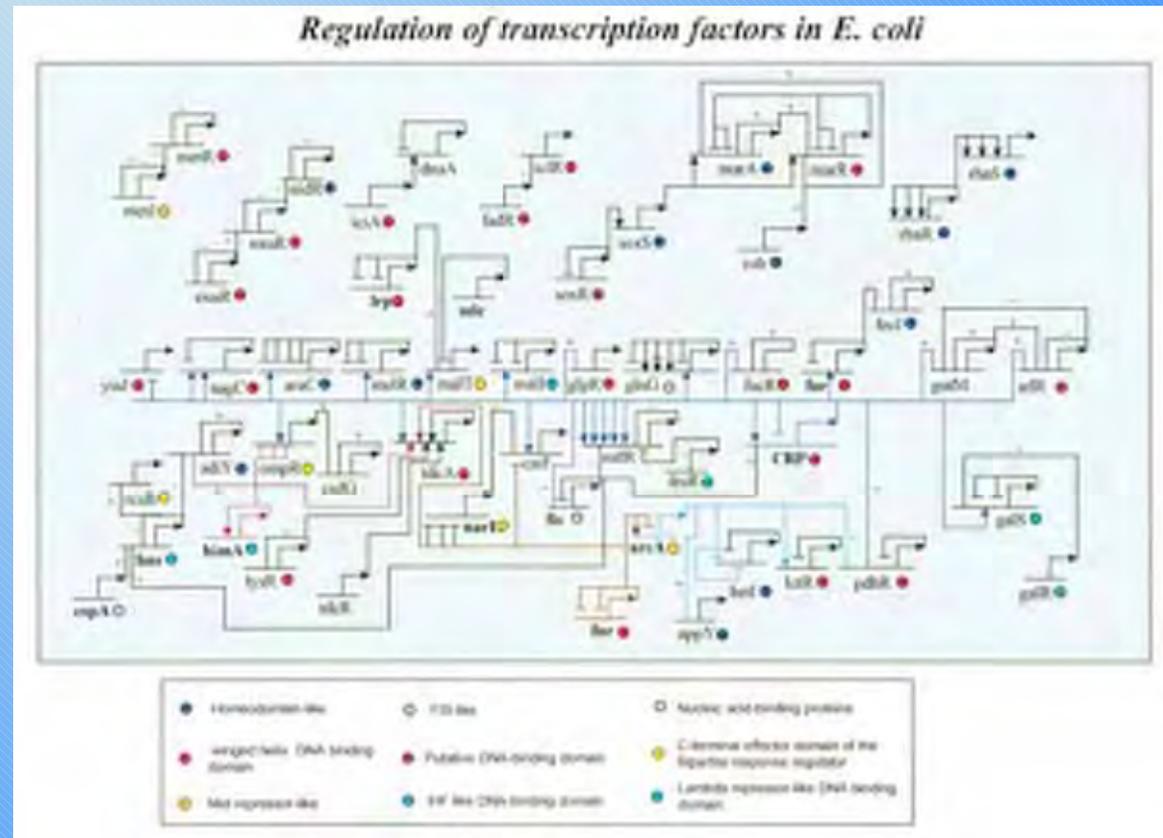


Interaction **network** (logical)





# Gene regulation in *Escherichia coli*



Logical network

M.M. Babu et al, Nucl. Acid Res. 31, 1234 (2003)



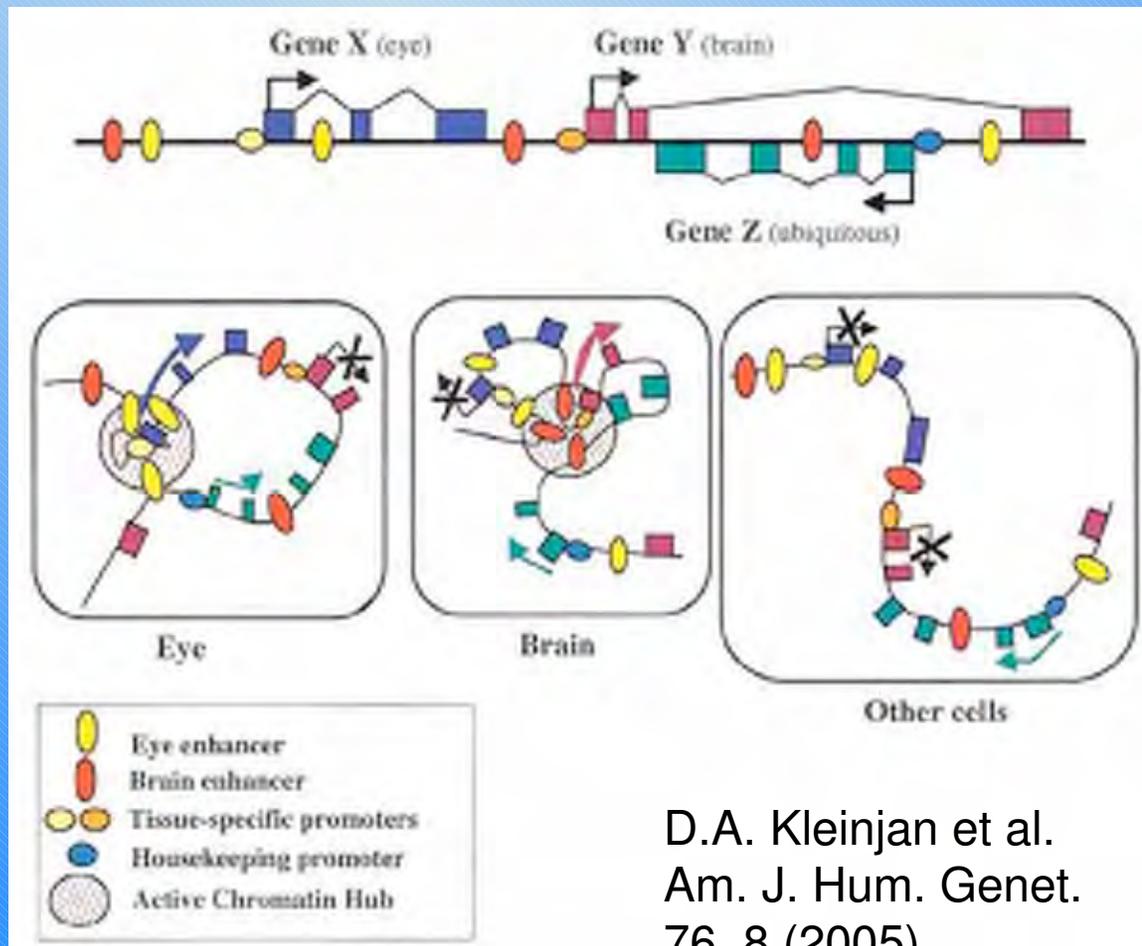
# What about multicellular organisms?

---



*Strongylocentrotus  
purpuratus*  
(Purple sea urchin)

# Genetic logic for multiple cells



D.A. Kleinjan et al.  
Am. J. Hum. Genet.  
76, 8 (2005)

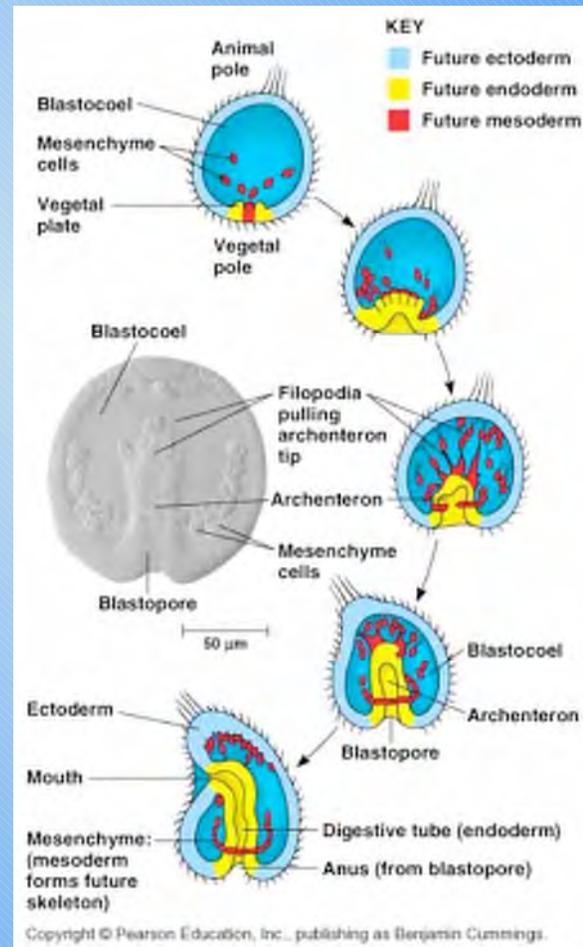
# Tissue differentiation: Phenotype expression

Development of the sea urchin embryo from a mass of undifferentiated cells (blastula)

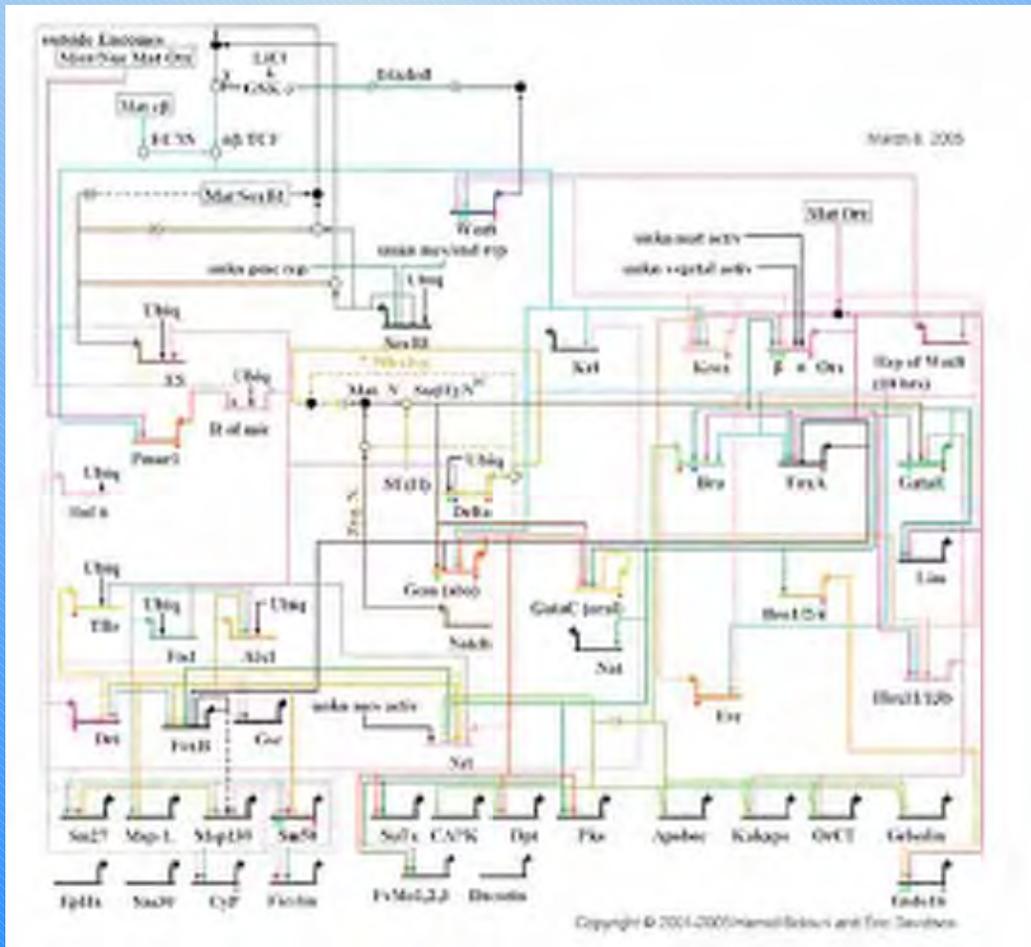
Each cell type has its own genetic regulatory cycle

The DNA must contain one possible cycle per cell type

More complex organisms =  
More complex DNA



# Genetic regulation during development

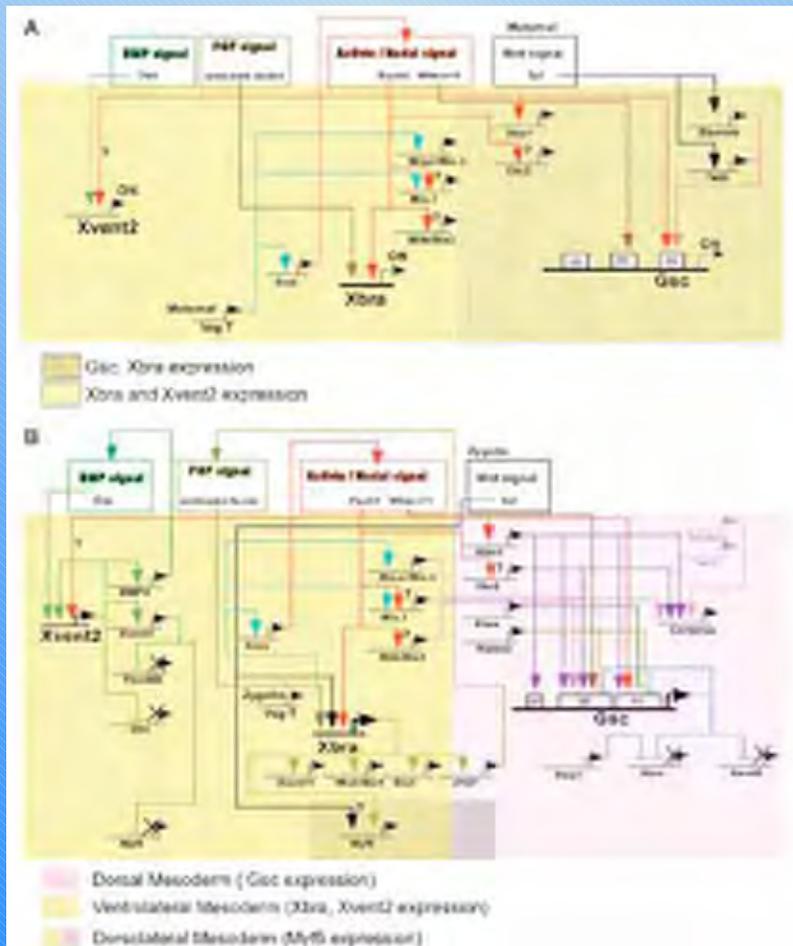


Endomesoderm  
genetic regulatory  
network for *S. Purpuratus*

W.J.R. Longabaugh et al,  
Dev. Biol. 283, 1 (2005)



# *Xenopus laevis* (African clawed frog)



Mesodermal genetic regulatory network  
 T. Koide et al, PNAS 102, 4943 (2005)



# Arabidopsis thaliana

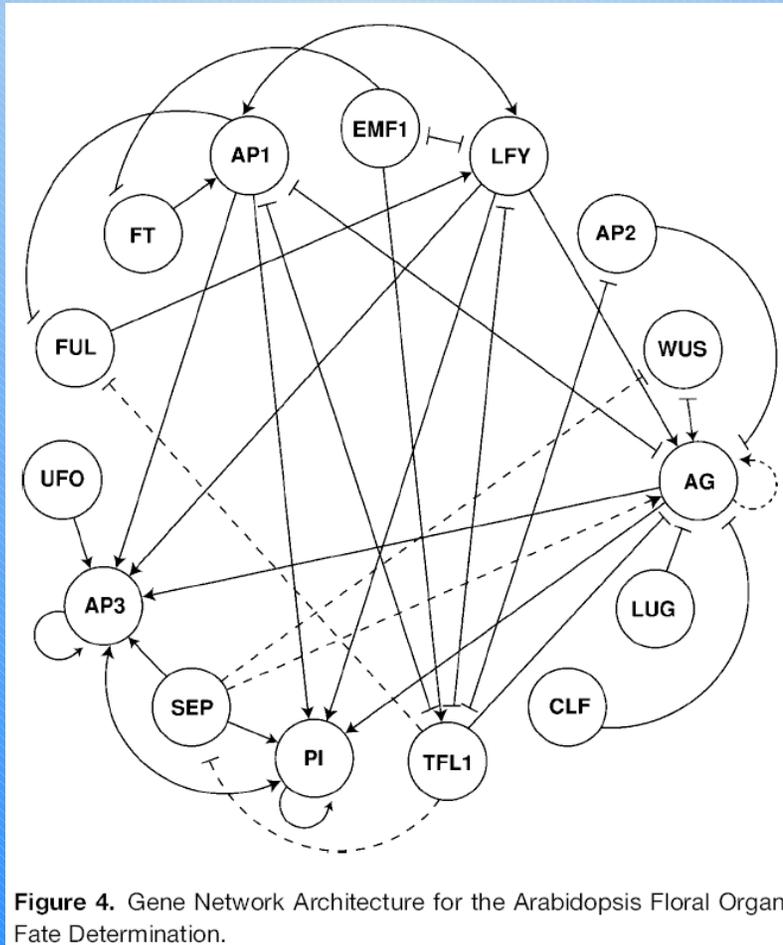


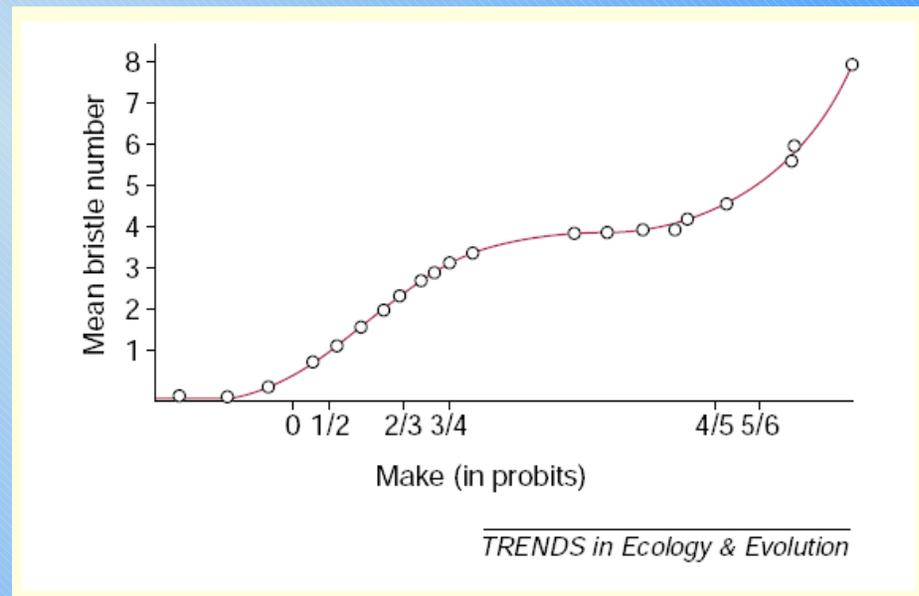
Figure 4. Gene Network Architecture for the Arabidopsis Floral Organ Fate Determination.



C. Espinosa-Soto et al, Plant Cell 16, 2923 (2004)



# Resistance to phenotype mutation



*Drosophila* sp.

J.M. Rendel, *Evolution* 13, 425 (1959).

C.D. Meiklejohn, D.L. Hartl, *Trends Ecol. Evol.*  
17, 468 (2002).

# Canalization: A biologist's view

---

C.H. Waddington, Nature 150, 563 (1942).

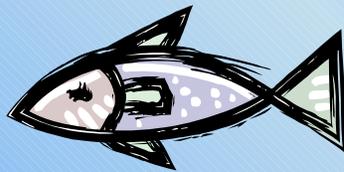
“Canalization and the inheritance of acquired characters”

“Once the developmental path has been canalized, it is to be expected that many different agents, including a number of mutations available in the germplasm of the species, will be able to switch development into it. By such a series of steps, then, it is possible that an adaptive response can be fixed without waiting for the occurrence of a mutation.”

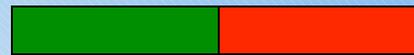


Canalization is mediated by “developmental reactions [that] are adjusted so as to bring about one definite end result regardless of minor variations in conditions during the course of the reaction.”

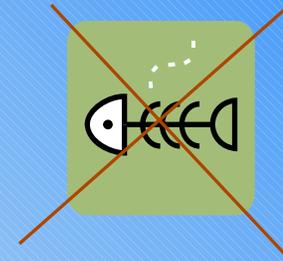
# Canalization: Allowing evolution to work offline?



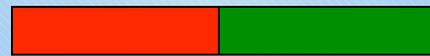
Fixed phenotype



Active      Inactive



Expression of new phenotype due to changing conditions



Inactive      Active

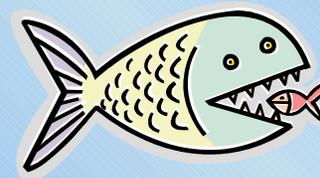
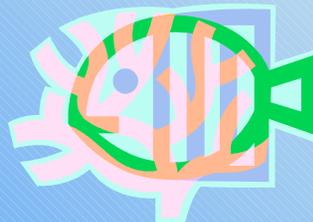
Mating display

(Lethal)

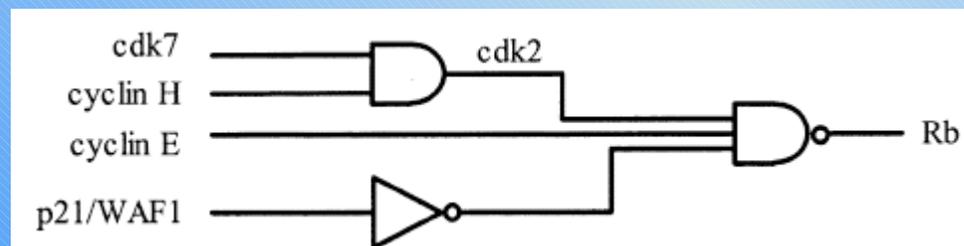
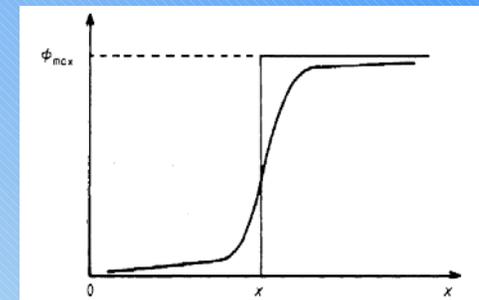
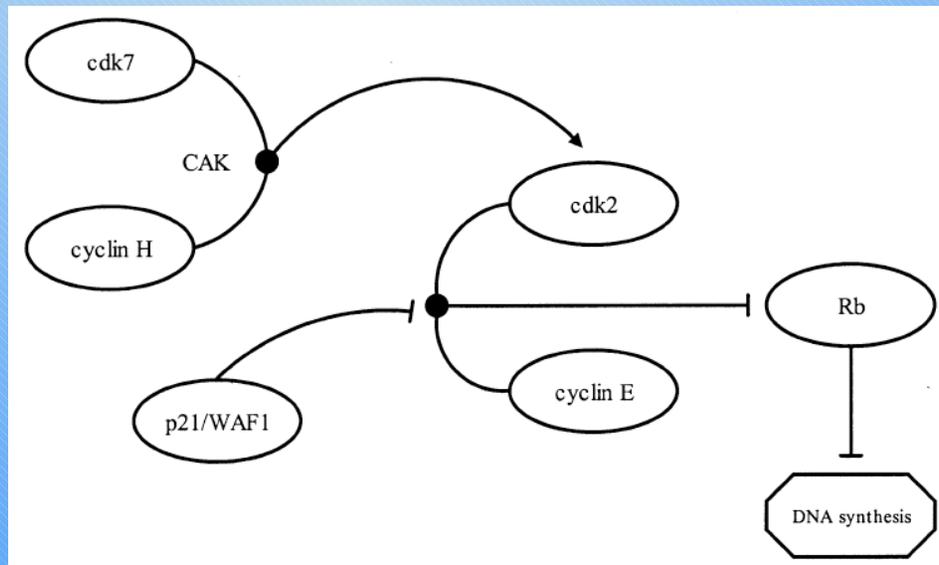
Camouflage

Better hunting

?



# Genetic interactions in the eyes of a physicist

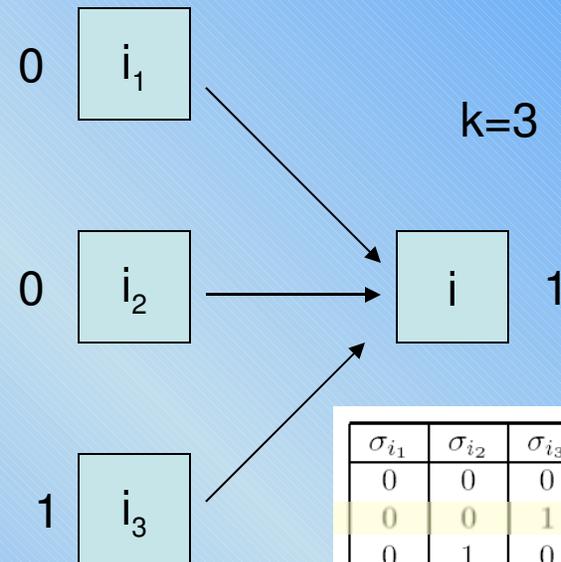


I. Shmulevich et al, Proc. IEEE 90, 1778 (2002)

# Kauffman's genetic logic model

S.A. Kauffman, "Metabolic stability and epigenesis in randomly constructed genetic nets," J. Theor. Biol. 22, 437 (1969)

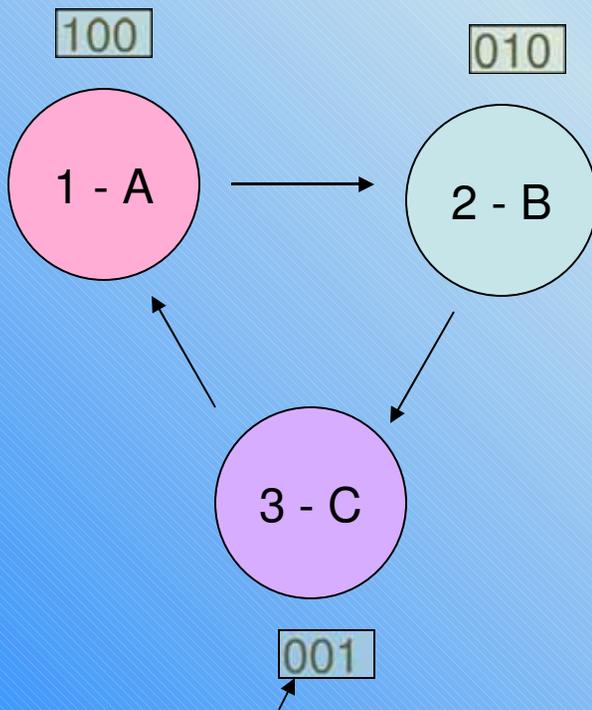
- Replace each gene with a Boolean logic element having two states, "off (0)" and "on (1)"
- Genes are randomly connected to  $k$  other input genes in a network
- The response of each gene to its  $k$  inputs is given by a randomly chosen Boolean strategy table



$\sigma_{i_1}$	$\sigma_{i_2}$	$\sigma_{i_3}$	$f_i$
0	0	0	0
0	0	1	1
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	0
1	1	1	0

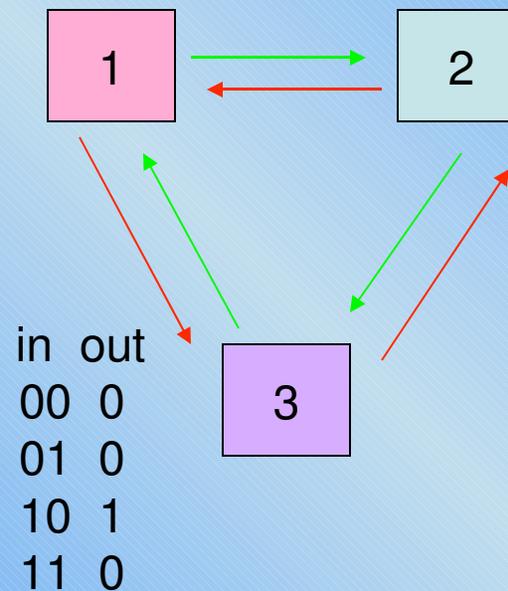
# Gene-gene interaction network

Cell **cycle** (temporal)



Current **state** of system  
100 -> 010 -> 001 -> 100

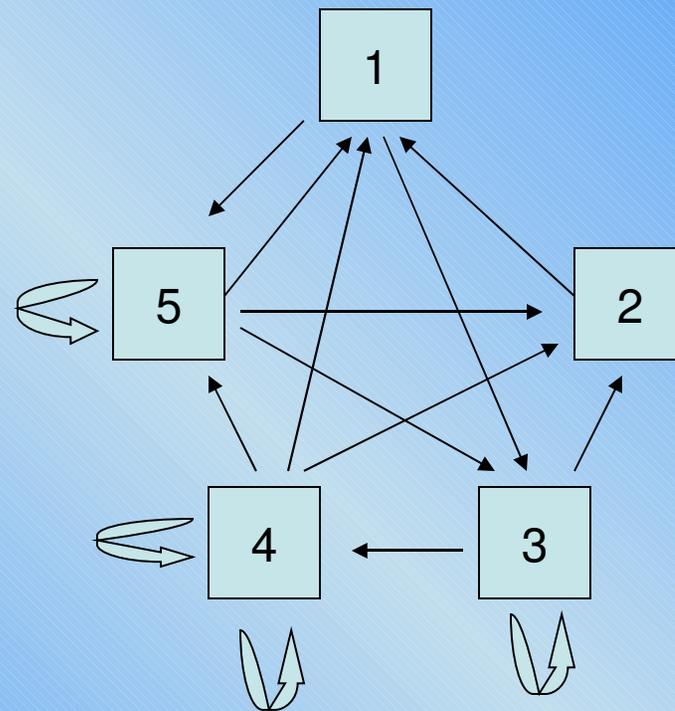
Interaction **network** (logical)



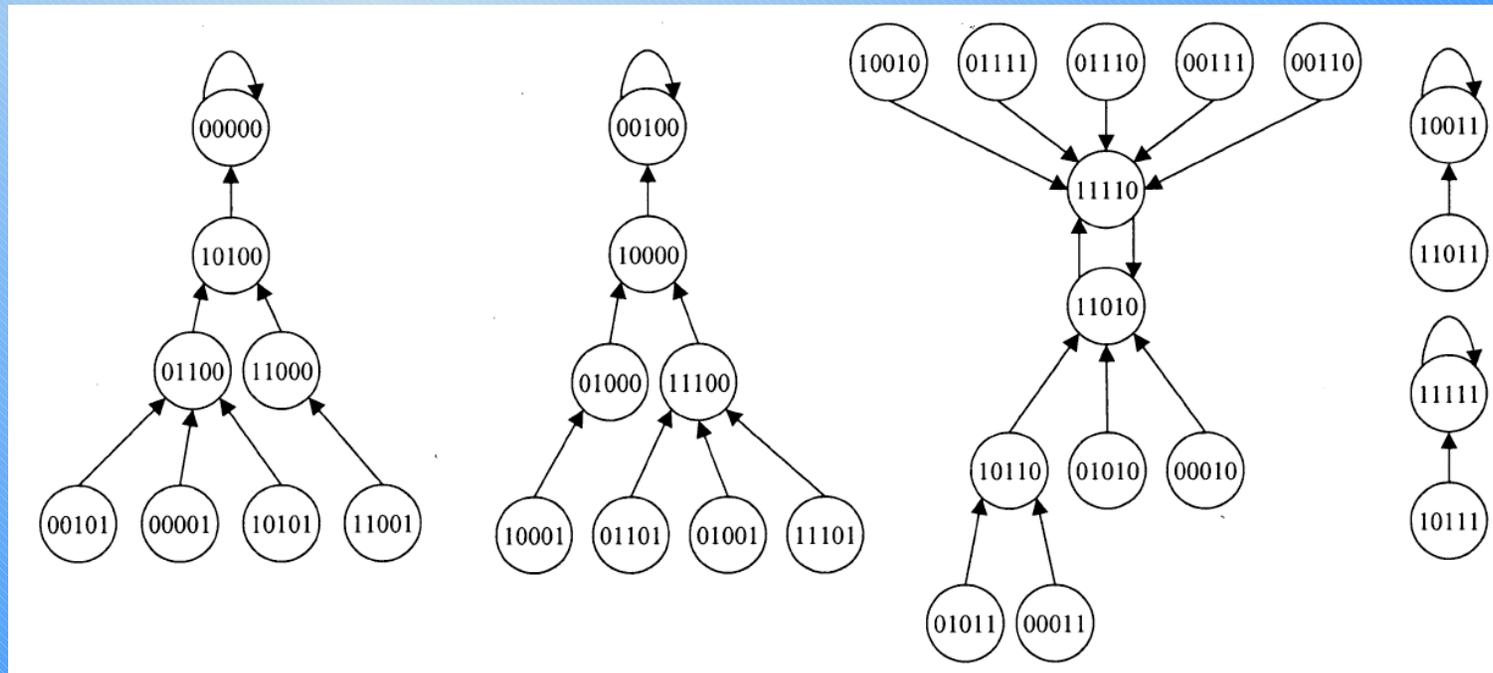
# Cycles in a random Boolean network

	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
	0	0	0	0	0
	1	1	1	0	0
	1	1	1	0	0
	1	0	0	1	0
	0	0	1	0	0
	1	1	1	1	0
	1	1	0	1	0
	1	1	1	1	1
$j_1$	5	3	3	3	5
$j_2$	2	5	1	4	4
$j_3$	4	4	5	4	1

$N=5, k=3$



# Cycles in a random Boolean network



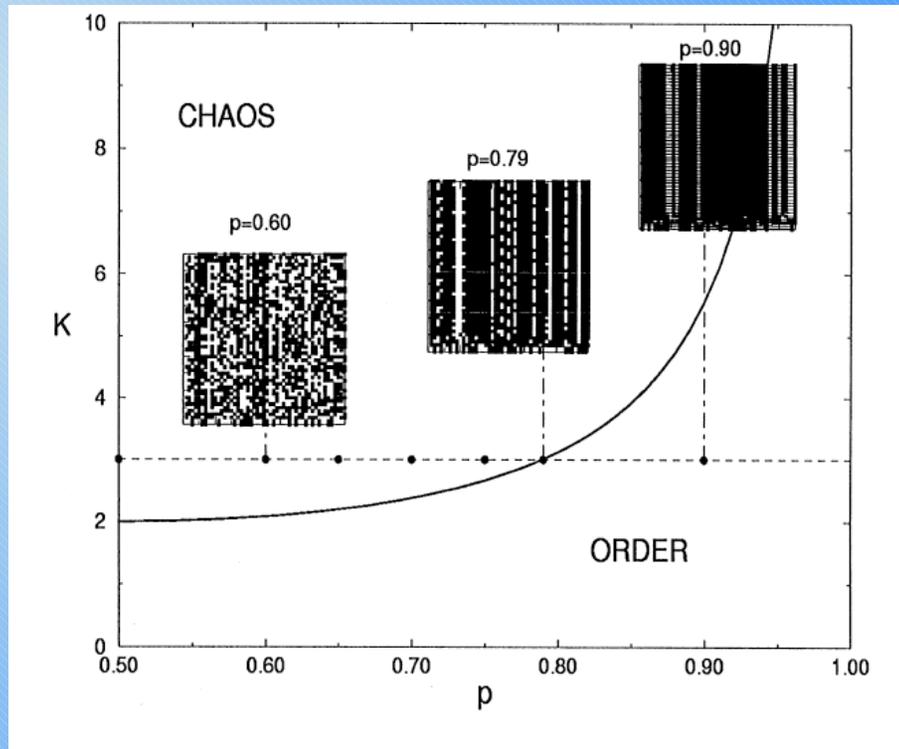
Kauffman: Each attractor corresponds to a **cell type**

I. Shmulevich et al, Proc. IEEE 90, 1778 (2002)

# Internal homogeneity

p=0.5	p=0.75
00 1	00 1
01 0	01 1
10 1	10 1
11 0	11 0

p=0.75	p=1.0
00 0	00 1
01 0	01 1
10 1	10 1
11 0	11 1

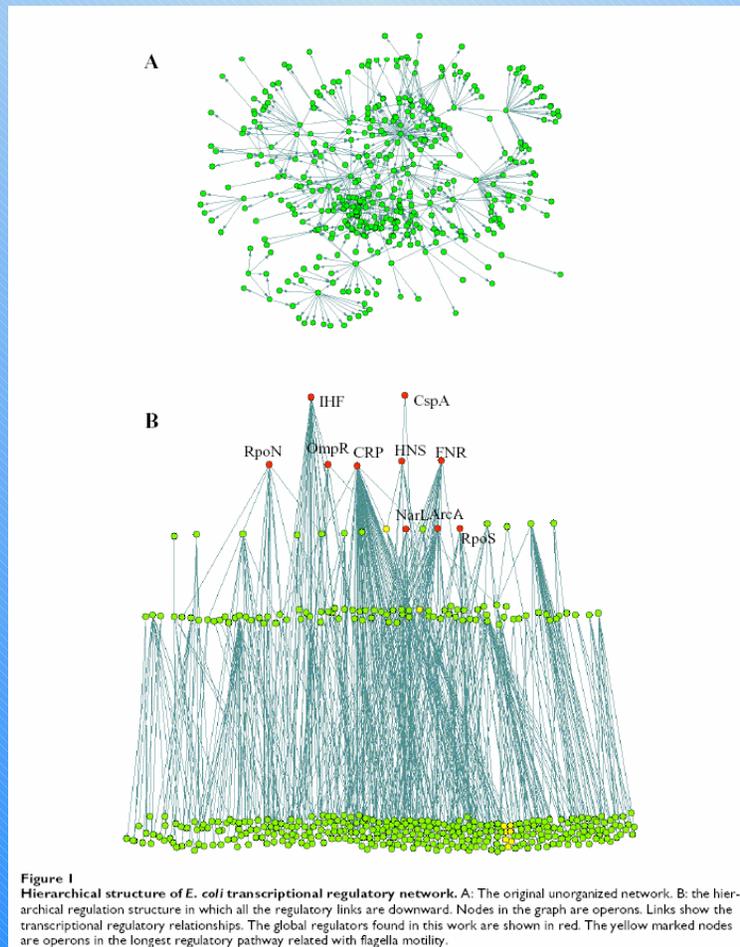


$$K = \frac{1}{2p(1-p)}$$

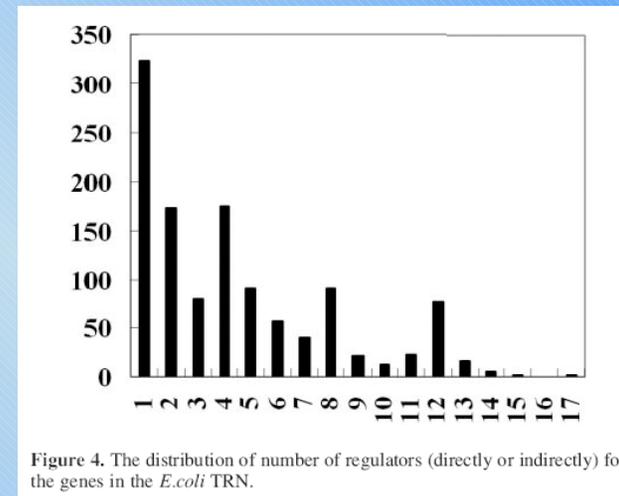
Avg period increases exponentially with N in chaotic regime; as a power of N in the ordered regime; intermediate at boundary.

B. Luque et al, Physica A 284, 33 (2000)

# Values of k in real genetic networks: *E. coli*

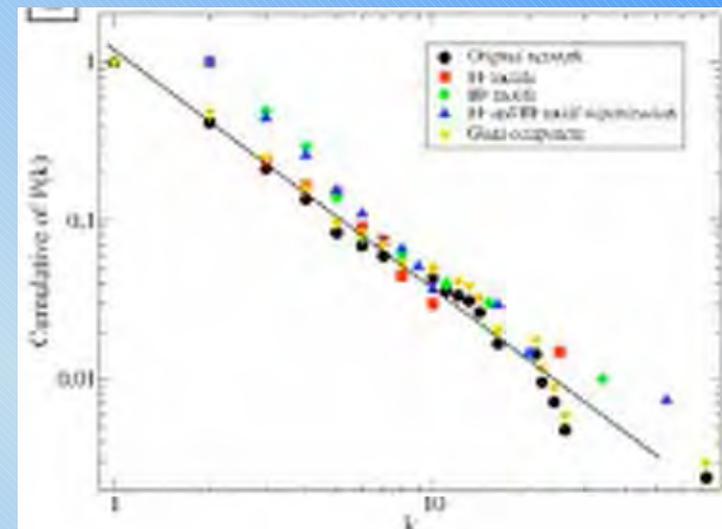
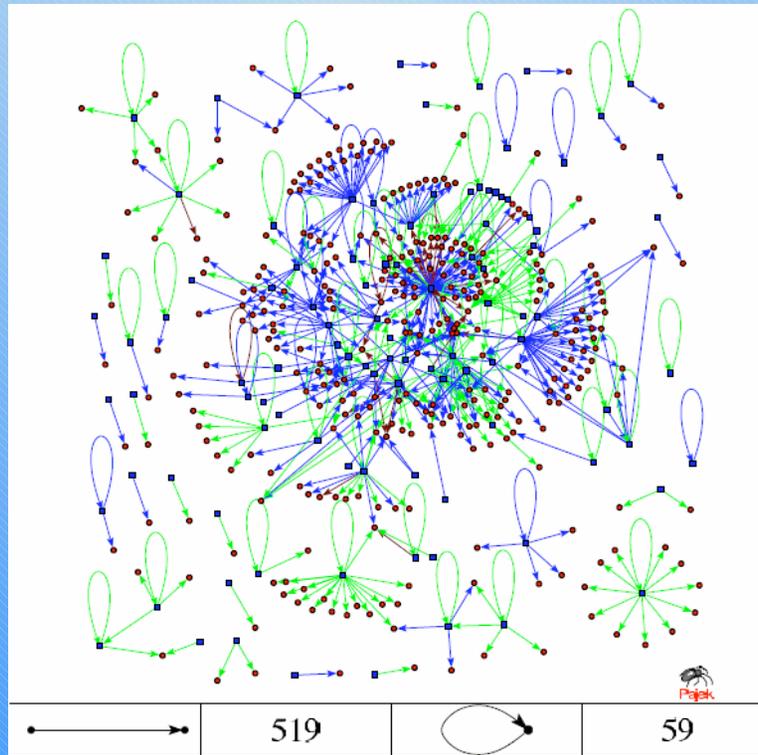


H.W. Ma et al, BMC Bioinformatics  
5, 199 (2004)



H.W. Ma et al, Nucl. Acids Res.  
32, 6643 (2004)

# Small values of $k$ and much autoregulation



Yeast transcription regulatory network structure  
R. Dobrin et al, BMC Bioinformatics 5, 10 (2004)

# Classifying strategies

---

Number of strategies:  $2^{2^k}$

k	$N_s$
2	2
3	16
4	256
5	65536
6	4294967296
7	18446744073709551616
8	Much larger than Avogadro's number

Due to combinatoric explosion, it is not possible to reach  $k=10$  by direct inspection

# Classifying k=2 strategies

TABLE I. The 16 update functions for nodes with 2 inputs. The first column lists the 4 possible states of the two inputs; the other columns represent one update function each, falling into four classes.

In	$\mathcal{F}$	$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{R}$
00	1 0	0 1 0 1	1 0 0 0 0 1 1 1	1 0
01	1 0	0 1 1 0	0 1 0 0 1 0 1 1	0 1
10	1 0	1 0 0 1	0 0 1 0 1 1 0 1	0 1
11	1 0	1 0 1 0	0 0 0 1 1 1 1 0	1 0

Fixed

Sensitive to  
only one input  
(*Acts like k=1*)

Sensitive to  
one state  
of one input

Sensitive to  
both inputs

p=1

p=0.5

p=0.75

p=0.5



V. Kaufman et al, PRE 72, 046124 (2005)



# Minority game for an evolving N-K network

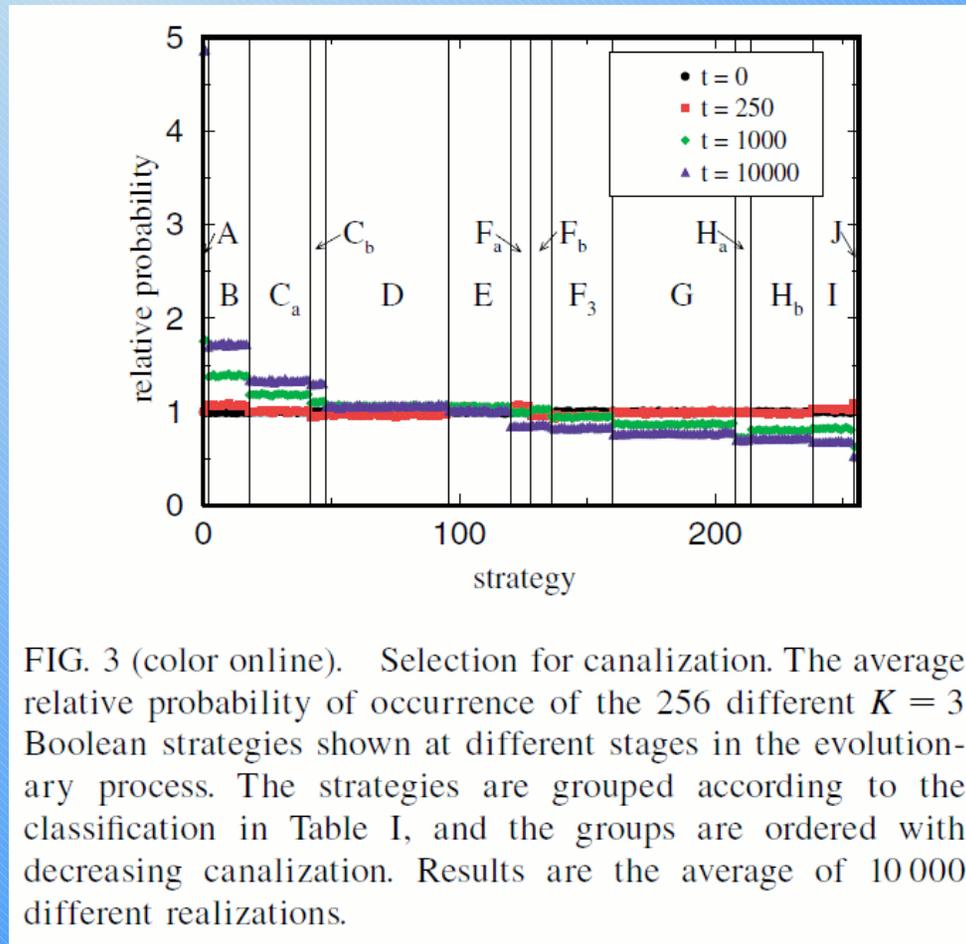
---

- System of  $N$  nodes is initialized with fixed  $K$ . All nodes are assigned an unbiased strategy (equal number of 0s and 1s).
- Repeatedly update the network until the attractor is reached.
- For each update on the attractor, determine whether “0” or “1” is the output of the majority of the nodes.
- Assign a “point” +1 to all nodes in the majority on each update.
- The node which is in the majority most often (has the most “points”) loses the game and is assigned a new randomly chosen unbiased strategy. This completes an “epoch.”
- Repeat the procedure for the new network.

K.E. Bassler, C. Lee, Y. Lee, PRL 93, 038101 (2004)



# Fourteen $k=3$ classes of strategies with equal evolutionary advantage



K.E. Bassler, C. Lee, Y. Lee, PRL 93, 038101 (2004)

# Classifying k=3 strategies

K.E. Bassler, C. Lee, Y. Lee, PRL 93, 038101 (2004)

Class	A	B	$C_a$	$C_b$	D	E	$F_a$	$F_b$	$F_c$	G	$H_a$	$H_b$	I	J
Size	2	16	24	6	48	24	8	8	24	48	6	24	16	2
$\mathcal{P}_0$	1	0												
$\mathcal{P}_1$	1	1/2	1/3	1/6			0							
$\mathcal{P}_2$	1	3/4	2/3	7/12	1/2	1/2		5/12	1/3	1/4	0			
$h$	1	7/8	3/4	1/2	5/8	3/4	3/4	1/2	5/8	1/2	5/8	1/2		
Sym	S	N	N	A	N	N	S	A	N	N	S	N	N	A

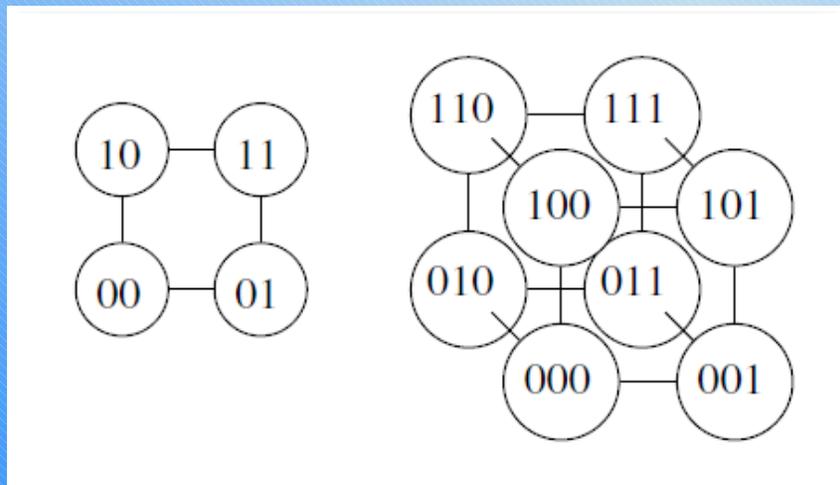
**Table 2.** Classification of the 256  $K = 3$  Boolean functions according to their canalization properties, internal homogeneity, and parity symmetry. (S indicates symmetric, A indicates anti-symmetric, and N indicates neither, or non-symmetric.)

Strategies in the same class have the same evolutionary advantage when the network is allowed to evolve under some rule.

How to identify classes? What about  $k > 3$ ?

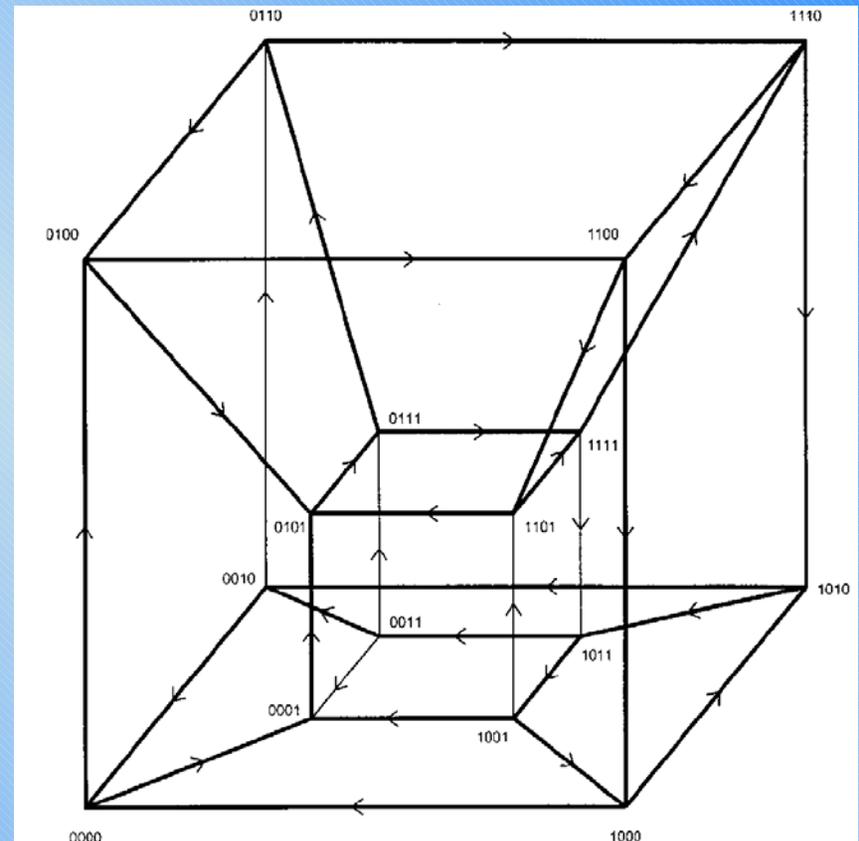


# Map strategies to Ising k-hypercubes



k=2

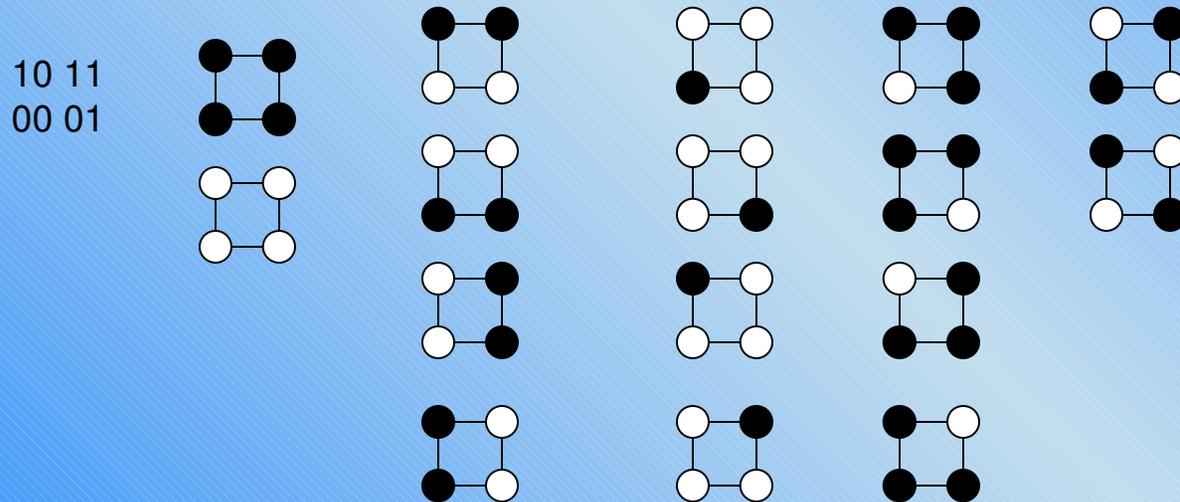
k=3



k=4

# Example: $k=2$

00	1 0	0 1 0 1	1 0 0 0 0 1 1 1	1 0
01	1 0	0 1 1 0	0 1 0 0 1 0 1 1	0 1
10	1 0	1 0 0 1	0 0 1 0 1 1 0 1	0 1
11	1 0	1 0 1 0	0 0 0 1 1 1 1 0	1 0



Each strategy group contains all objects in a particular group orbit of the Ising hypercube symmetry group plus parity

# Counting strategy classes

---

All strategies correspond to all states of the k-hypercube.

We can identify the total number of strategy classes by counting the number of group orbits that exist for the k-hypercube.

$N_c$	k	$N_s$
-	2	2
4	3	16
14	4	256
	5	65536
	6	4294967296
	7	18446744073709551616

# Permutations and Cyclic Decomposition

---

Given an ordered set of elements, a permutation is a reordering of that set where each element occurs only once.

“game”   “emag”   “ameg”   “eagm”  
{1,2,3,4}   {4,3,2,1}   {2,3,4,1}   {4,2,1,3}

Cyclic decomposition: Consider the permutation {4,2,1,3} of {1,2,3,4}. Repeated applications of this permutation result in a cycle:

game -> eagm -> maeg -> game

The permutation can be written in terms of cycles of elements:  
(2)(143)

# Basic Group Theory

---

**Definition 1.1.** A *group*  $(G, \cdot)$  is a set  $G$  with a binary operation

$$\cdot : G \times G \rightarrow G,$$

and a unit  $e \in G$ , possessing the following properties.

- (1) Unital: for  $g \in G$ , we have  $g \cdot e = e \cdot g = g$ .
- (2) Associative: for  $g_i \in G$ , we have  $(g_1 \cdot g_2) \cdot g_3 = g_1 \cdot (g_2 \cdot g_3)$ .
- (3) Inverses: for  $g \in G$ , there exists  $g^{-1} \in G$  so that  $g \cdot g^{-1} = g^{-1} \cdot g = e$ .

A set of elements  $S$  of  $G$  is said to *generate*  $G$  if every element of  $G$  may be expressed as a product of elements of  $S$ , and inverses of elements of  $S$ . That is to say, given  $g \in G$ , there exists  $s_i \in S$  and  $\epsilon_i \in \{\pm 1\}$  so that

$$g = s_1^{\epsilon_1} \cdots s_n^{\epsilon_n}.$$

If a group  $G$  is generated by a single element, it is said to be *cyclic*. Every element of a cyclic group  $G$  is of the form  $g^n$  for some  $n \in \mathbb{Z}$ .

# Orbit-counting theorem

---

Total number of classes  $P_G$

$$P_G(x_1, x_2, \dots) = \frac{1}{|G|} \sum_{g \in G} |X^g|$$

$|G|$ : number of symmetry operators (generators)  $g$

$X^g$ : the set of elements in  $X$  that are left invariant by  $g$

Counting the number of classes for higher  $k$ :

- Identify the symmetry operators of the  $k$ -hypercube with parity
- Write these symmetry operators in terms of cycles
- Find the number of fixed points for each symmetry operator

The symmetry group for the  $k$ -hypercube is isomorphic to the hyperoctahedral group  $O_n$  with  $n=k$ , which has  $n!2^n$  symmetry operations

## Example: $k=2$

---

- Identify the symmetry operators of the  $k$ -hypercube with parity
- Write these symmetry operators in terms of cycles
- Find the number of fixed points for each symmetry operator

$$k!2^k = 8 \text{ for } k=2$$

(1)(2)(3)(4) E	$2^4$	(1)(2)(3)(4) P	0
(1243) E	2	(1243) P	2
(3421) E	2	(3421) P	2
(14)(23) E	$2^2$	(14)(23) P	$2^2$
(12)(34) E	$2^2$	(12)(34) P	$2^2$
(13)(24) E	$2^2$	(13)(24) P	$2^2$
(14)(2)(3) E	$2^3$	(14)(2)(3) P	0
(23)(1)(4) E	$2^3$	(23)(1)(4) P	0

$$48/16 = 4 \text{ classes}$$

Identifying generators by inspection is difficult for  $k>3$



# Arbitrarily high k: Cycle representation of the Zyklenzeiger group

THEOREM 7. The cycle index of  $\mathcal{G}_n$  is given by

$$Z_{\mathcal{G}_n} = \frac{1}{n!2^n} \sum_{(j)} \frac{n! 2^n}{\prod_{i=1}^n j_i! (2i)^{j_i}} \prod_{i=1}^n \left( \prod_{\substack{d|1 \\ d \neq 1}} f_d^{\epsilon(d)} + \prod_{\substack{d|2i \\ d \neq 1}} f_d^{g(d)} \right)^{j_i}$$

where the sum is over all solutions of

$$\sum_{i=1}^n ij_i = n;$$

where

$$\epsilon(k) = \frac{1}{k} \sum_{d|k} 2^d \mu\left(\frac{k}{d}\right)$$

and

$$g(2k) = \frac{1}{2k} \sum_{\substack{d|2k \\ d \neq 1}} 2^{d/2} \mu\left(\frac{2k}{d}\right)$$

where  $\mu(a)$  is the Möbius function.

# Example: k=2

$$x_1^4 + 3x_2^2 + 2x_1^2x_2 + 2x_4$$

$x_1^4$	(1)(2)(3)(4) E	$2^4$	(1)(2)(3)(4) P	0
$x_4$	(1243) E	2	(1243) P	2
$x_4$	(3421) E	2	(3421) P	2
$x_2^2$	(14)(23) E	$2^2$	(14)(23) P	$2^2$
$x_2^2$	(12)(34) E	$2^2$	(12)(34) P	$2^2$
$x_2^2$	(13)(24) E	$2^2$	(13)(24) P	$2^2$
$x_2^2$	(14)(2)(3) E	$2^3$	(14)(2)(3) P	0
$x_1^2x_2$	(23)(1)(4) E	$2^3$	(23)(1)(4) P	0

For each operator without parity, the number of functions left invariant is equal to  $2^{N_c}$ , where  $N_c = \sum_{i=1}^k b_i$  is the total number of cycles in the operator. Parity must be treated separately; no functions are left invariant by the parity operator with any  $k$ -hypercube operator containing at least one cycle of length 1. Thus there are  $2^{N_p}$  functions left invariant for the eight operators which include parity, where  $N_p = (1 - \Theta(b_1)) \sum_{i=1}^k b_i$  and  $\Theta$  is the Heaviside step function.

# Generating polynomials through k=5

**Table 3.** Cycle polynomials for  $k = 1$  through 5 and the number of classes  $P_G$  for each  $k$ .

$k$	Cycle polynomial	$P_G$
1	$(1/2) (x_1^2 + x_2)$	2
2	$(1/8) (x_1^4 + 3x_2^2 + 2x_1^2x_2 + 2x_4)$	4
3	$(1/48) (x_1^8 + 13x_2^4 + 8x_1^2x_3^2 + 8x_2x_6 + 6x_1^4x_2^2 + 12x_4^2)$	14
4	$(1/384) (x_1^{16} + 12x_1^8x_2^4 + 51x_2^8 + 12x_1^4x_2^6 + 32x_1^4x_3^4 + 48x_1^2x_2x_4^3 + 84x_4^4 + 96x_2^2x_6^2 + 48x_8^2)$	238
5	$(1/3840) (x_1^{32} + 384x_{10}^3x_2 + 20x_1^{16}x_2^8 + 60x_1^8x_2^{12} + 231x_2^{16} + 80x_1^8x_3^8 + 320x_{12}^2x_4^2 + 240x_1^4x_2^2x_4^6 + 240x_2^4x_4^6 + 520x_4^8 + 384x_1^2x_5^6 + 160x_1^4x_2^2x_3^4x_6^2 + 720x_2^4x_4^4 + 480x_8^4)$	698 635

# Class structure: Isomer chemistry

Substitute a term of the form  $A^a B^b$  for each  $x_a$

**Table 4.** Class structure for  $k = 2$ .

Class type	$N_h$	$\langle S_c \rangle$
$A^4$	1	2
$A^3 B$	1	8
$A^2 B^2$	2	3

$k=2$  (max 16)

**Table 5.** Class structure for  $k = 3$ .

Class type	$N_h$	$\langle S_c \rangle$
$A^8$	1	2
$A^7 B$	1	16
$A^6 B^2$	3	18.667
$A^5 B^3$	3	37.333
$A^4 B^4$	6	11.667

$K=3$  (max 96)

**Table 6.** Class structure for  $k = 4$ .

Class type	$N_h$	$\langle S_c \rangle$
$A^{16}$	1	2
$A^{15} B$	1	16
$A^{14} B^2$	4	60
$A^{13} B^3$	6	186.667
$A^{12} B^4$	19	191.58
$A^{11} B^5$	27	323.56
$A^{10} B^6$	50	320.32
$A^9 B^7$	56	408.57
$A^8 B^8$	74	173.9

$K=4$  (max 768)

$$S_c^{\max} = k! 2^{k+1}$$

$$\langle S_c \rangle = N_f / N_h$$

$$N_f(m, n) = (2 - \delta_{m,n})(2^k)! / (m!n!), \text{ where } m + n = 2^k,$$



# Class structure for $k=5$

**Table 7.** Class structure for  $k = 5$ .

Class type	$N_h$	$\langle S_c \rangle$
$A^{32}$	1	2
$A^{31}B$	1	64
$A^{30}B^2$	5	198.4
$A^{29}B^3$	10	992
$A^{28}B^4$	47	1530.2
$A^{27}B^5$	131	3074.4
$A^{26}B^6$	472	3839.8
$A^{25}B^7$	1326	5076.7
$A^{24}B^8$	3779	5566.7
$A^{23}B^9$	9013	6224.1
$A^{22}B^{10}$	19963	6463.2
$A^{21}B^{11}$	38073	6777.7
$A^{20}B^{12}$	65664	6877.2
$A^{19}B^{13}$	98804	7031.6
$A^{18}B^{14}$	133576	7058.7
$A^{17}B^{15}$	158658	7131.3
$A^{16}B^{16}$	169112	3554.3

$K=5$  (7680)



# Characteristic polynomials

---

Do two randomly chosen strategies belong to the same class?

Each class has a unique characteristic polynomial.

Construct the adjacency matrix:  $A_{ij}=1$  if link, 0 if no link

Place the strategy on the diagonal

Find the determinant

a 1 1 0

1 b 0 1

1 0 c 1

0 1 1 d

Det:  $-ab-ac-bd-cd-abcd$

$-4A^2+A^4$

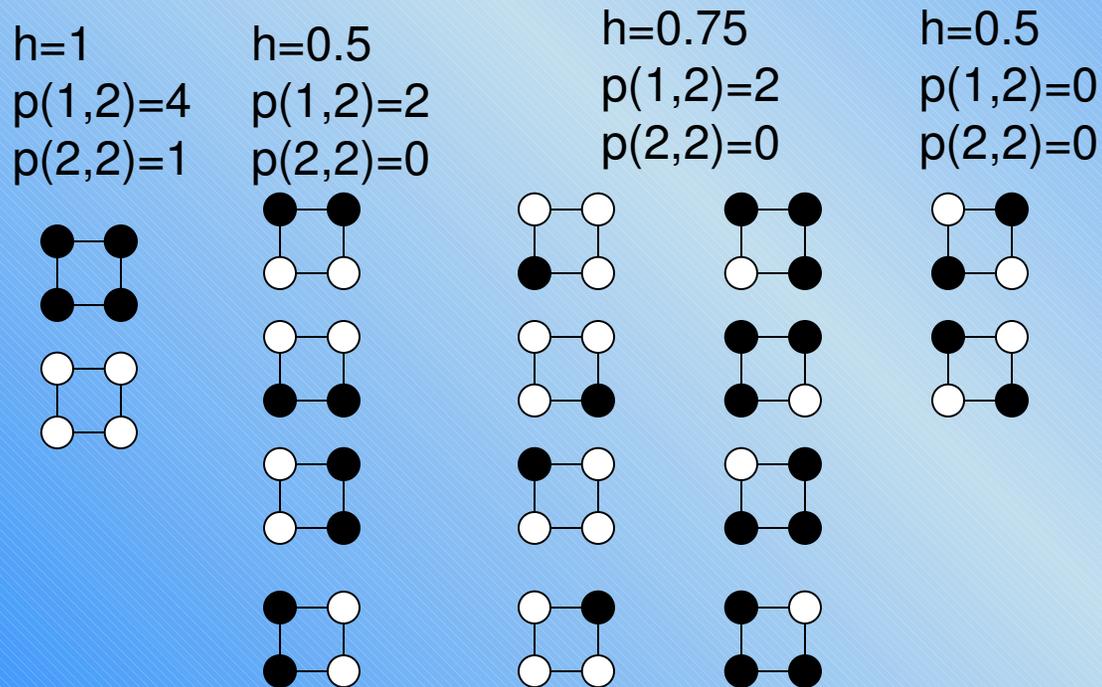
$-4AB+A^2B^2$

$-A^2-2AB-B^2+A^2B^2$

$-2AB-2B^2+A^2B^2$

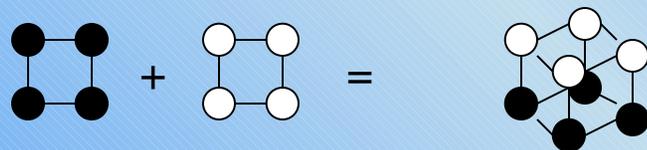
# Geometric strategy classification

We classify a given strategy depending on the number of edges, faces, and higher dimensional objects having all entries the same,  $p(d,k)$ ,  $d \leq k$ . This represents varying degrees of canalization.



# Recursion relations

We can think of a  $k+1$  strategy as being assembled out of two  $k$  strategies



00 1	00 0	000 1
01 1	01 0	001 1
10 1	10 0	010 1
11 1	11 0	011 1
		100 0
$p(1,2)=4$	$p(1,2)=4$	101 0
$p(2,2)=1$	$p(2,2)=1$	110 0
		111 0
		$p(1,3)=8$
		$p(2,3)=2$
		$p(3,3)=0$

$$h(k+1) = \sum_{i=1}^2 h_i(k)$$

$$p(d, k+1) \geq \sum_{i=1}^2 p_i(d, k)$$

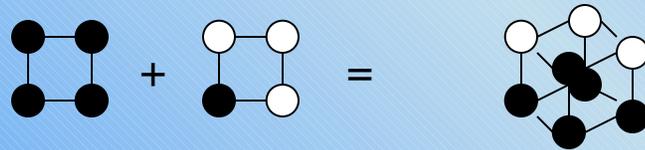
$$N_d(k) = \frac{2^{k-d} k!}{(k-d)! d!}$$

$$N_d(k+1) \geq p(d, k+1) \geq \sum_{i=1}^2 p_i(d, k)$$

$$\frac{N_d(k)}{N_d(k-1)} = \frac{2k}{k-d}$$

# Recursion relations

Improving on the maximum bound



00 1  
01 1  
10 1  
11 1

$p(1,2)=4$   
 $p(2,2)=1$

00 1  
01 0  
10 0  
11 0

$p(1,2)=2$   
 $p(2,2)=0$

000 1  
001 1  
010 1  
011 1  
100 1  
101 0  
110 0  
111 0

$p(1,3)=7$   
 $p(2,3)=1$   
 $p(3,3)=0$

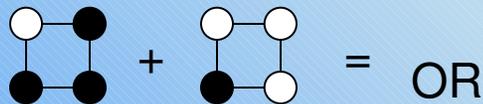
$$p(d, k + 1) \geq \sum_{i=1}^2 p_i(d, k)$$

$$p(d, k + 1) \leq \sum_{i=1}^2 p_i(d, k) + \min(p_1(d, k), p_2(d, k))$$

$$6 \leq p(1,3) \leq 8$$

# Lower symmetry cases

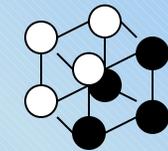
$$4 \leq p(1,3) \leq 6$$



00 1	00 1
01 1	01 0
10 0	10 0
11 1	11 0

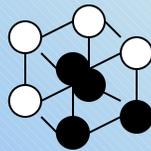
$p(1,2)=2$	$p(1,2)=2$
$p(2,2)=0$	$p(2,2)=0$

OR



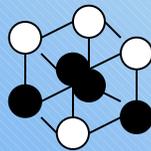
+ all possible rotations

$p(1,3)=6$   
 $p(2,3)=0$   
 $p(3,3)=0$   
 symmetric



+ all possible rotations

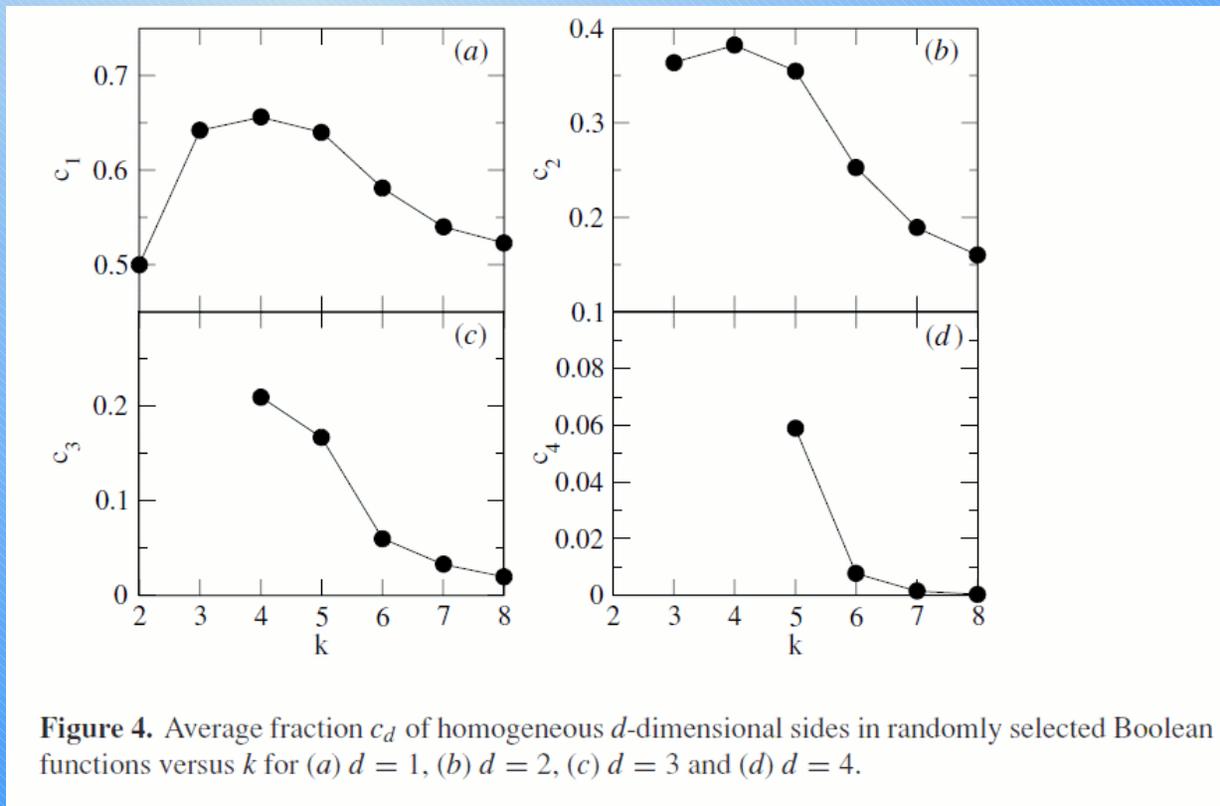
$p(1,3)=6$   
 $p(2,3)=0$   
 $p(3,3)=0$   
 asymmetric



+ all possible rotations

$p(1,3)=5$   
 $p(2,3)=0$   
 $p(3,3)=0$

# Bounding canalization



Although the fraction of fully canalizing functions drops rapidly with  $k$ , the fraction of partially canalizing functions remains large.

# Effective k may be lower than actual k

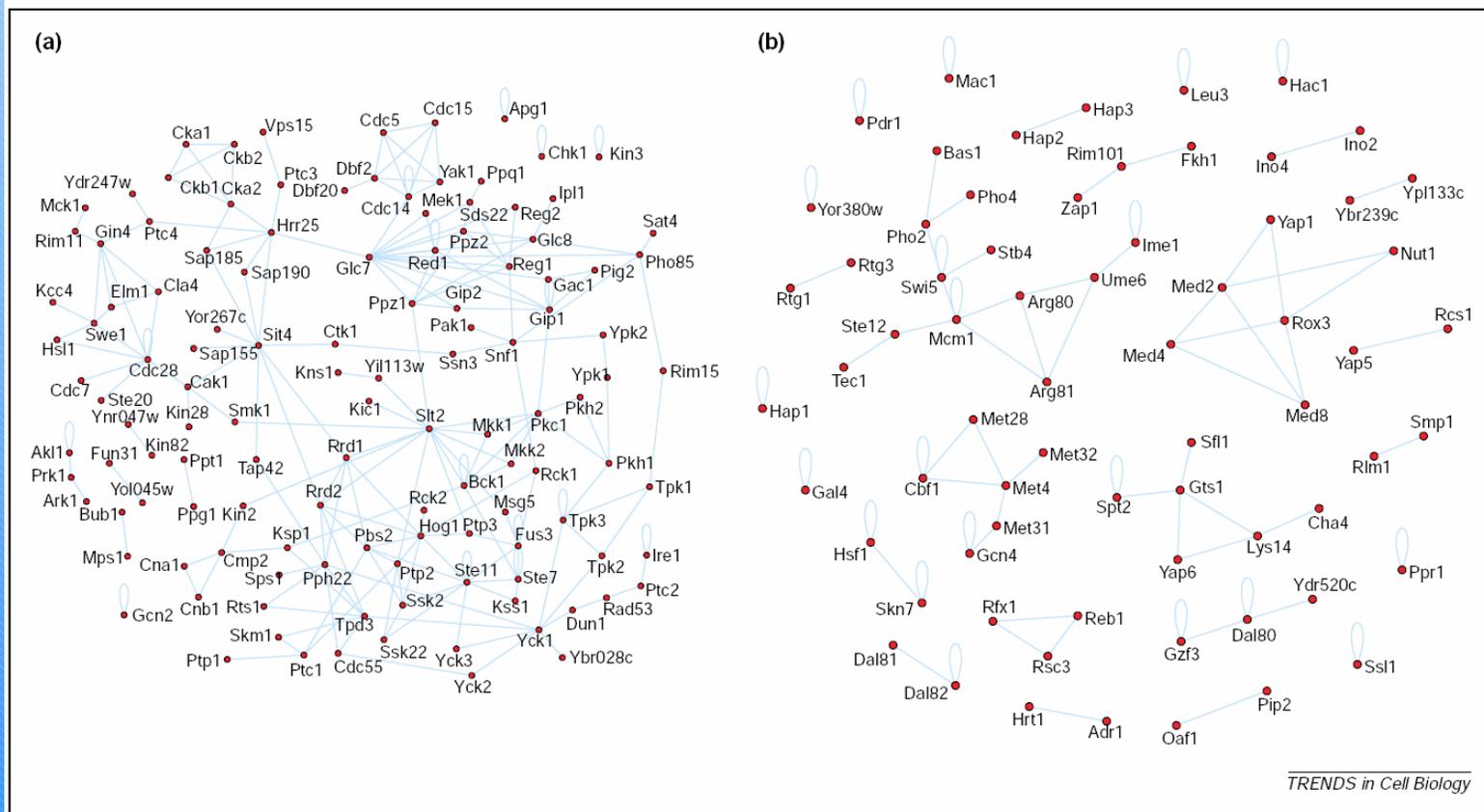


Fig. 5. A large network of protein-protein interactions among kinases and phosphatases in yeast. (a) Kinases and phosphatases are very well connected in a large protein-protein interaction network. (b) Transcription factors, a functional class similar in size to the kinase/phosphatase class, are not. This is an example of an unanticipated result that is completely unobtainable without genome-wide studies. Loops indicate self-interactions.

# Summary

---

- Boolean network models can be used to represent genetic regulatory networks.
- The number of possible strategies grows rapidly with connection degree  $k$ ; the number of network states grows rapidly with network size  $N$ .
- Nodes which respond to only a fraction of their inputs have an effectively reduced  $k$ , which reduces the available phase space.
- Mapping of the gene strategy tables to Ising hypercubes allows us to use symmetry properties to enumerate strategy classes.
- By assembling  $k+1$  strategies out of  $k$  strategies recursively, we can put bounds on the amount of canalization present in the  $k+1$  strategies.
- A significant fraction of strategies are at least partially canalized, reducing the complexity and cycle length of the logical network.